Building Sustainable Open Source AI Ecosystem

Ibrahim Haddad, Ph.D.
Executive Director, LF AI Foundation
Ibrahim@LinuxFoundation.org
Agenda

› The Linux Foundation and sustainable open source

› Challenges in the current open source AI landscape

› LF AI: Mission, structure, members, projects, ongoing efforts
The Linux Foundation and Sustainable Open Source
Today the Linux Foundation is much more than Linux

**Security**
We are helping global privacy and security through a program to encrypt the entire internet.

**Networking**
We are creating ecosystems around networking to improve agility in the evolving software-defined datacenter.

**Cloud**
We are creating a portability layer for the cloud, driving de facto standards and developing the orchestration layer for all clouds.

**Web**
We are providing the application development framework for next generation web, mobile, serverless, and IoT applications.

**Blockchain**
We are creating a permanent, secure distributed ledger that makes it easier to create cost-efficient, decentralized business networks.

**Automotive**
We are creating the platform for infotainment in the auto industry that can be expanded into instrument clusters and telematics systems.

**AI**
We are creating a sustainable open source AI ecosystem that makes it easier to create AI products and services using open technologies.
The Linux Foundation continues to grow

- **1400+** Members From 41 Countries
- **80%** of Fortune 100 Tech & Telecom
- **35,000+** Developers Contributing Code
- **170+** Open Source Projects
- **$16B** Shared Value

We have seen unprecedented growth in our projects
Linux Foundation project types (examples)
Open source isn’t slowing down anytime soon
The real question is: Which projects really matter?
LF seeks to accelerate new projects to adoption and sustainability

![Graph showing the relationship between the value of individual projects and the number of open source projects, with a focus on building ecosystems for critical and high-value projects.](image-url)
The answer: projects with sustainable ecosystems

Sustainable projects have a developer community
(1) whose technology is used in commercial solutions,
(2) that profit businesses,
(3) who in turn participate and reinvest back into the project and hire developers to work in the community
Building successful open collaboration ecosystems usually comes down to 4 core factors

1. Neutrality
   › No one company or organization can “take it away” from the community that forms around a project

2. Open Governance Do-ocracy
   › The most successful projects that have stood the test of time have neutral, open governance models where those who do the work make the decisions in a defined governance model

3. IP Clarity
   › Removal of IP uncertainty enables anyone to get involved as a contributor or implement as a user

4. Commercial Support Ecosystem
   › Encouraging commercial engagement in the project leads to jobs, faster adoption, new contributions and features that address new use cases
Creating sustainable OSS projects requires hard work

| Governance and Membership | • Incorporation, Tax status, Bylaws, Member Agreements, Antitrust, etc  
  • Ongoing business development and membership recruitment |
|---------------------------|------------------------------------------------------------------------|
| Development Process       | • Technical Decision Making  
  • Project Life Cycle  
  • Release Process |
| Infrastructure            | • Custom infrastructure using open source best practices  
  • Security and reliability |
| Ecosystem Development     | • Marketing  
  • Events  
  • Training |
| IP Management             | • Code Provenance and License compliance  
  • Trademark management  
  • Legal defense and Collaboration |
Challenges in the Current Open Source AI Landscape
AI is dominated by open source software!

160
Projects
965,000+ GitHub Stars

9
Categories

67
Organizations
Market cap of $6.49T

11
Universities

10
Open Source Licenses
You are currently tracking **355,762,809 lines of code**, committed by **30,204 developers**, from **282 known organizations**, working in **240 repos**, on **178 projects** over the last **5 years, 7 months, and 14 days**.
Challenges, Summary

Fragmentation, lack of integration and harmonization, lack of projects’ cross pollination

Developed for specific product needs. The open source spinoff is a result of wanting to build an ecosystem and collaborate on building a platform

no formal governance, governance favors the creators of the project, or project backer dominating dev and governance

Glass ceiling for project investment and adoption in the absence of a fair governance

As project grows, it is unclear who should own legal and administrative tasks that are essential to the health of the project
The promise of AI open source ecosystem

Unique reasons to AI that favor the open source model and ecosystem

**Fairness**
Methods to detect and mitigate bias in datasets and models, including bias against known protected populations

**Robustness**
Methods to detect alterations/tampering with datasets and models, including alterations from known adversarial attacks

**Explainability**
Methods to enhance understandability/interpretability by persona/roles in process of AI model outcomes/decision recommendations, including ranking and debating results/decision options

**Lineage**
Methods to ensure provenance of datasets and AI models, including reproducibility of generated datasets and AI models

*(Open) Data* - Requires cleaning, sorting, tagging and provenance tracking, and a governance structure for doing these things
The LF AI Foundation

Web site: https://lfdl.io
Landscape: http://l.lfdl.io
GitHub: https://github.com/LFDLFoundation
Wiki: https://wiki.deeplearningfoundation.org
What is the LF AI Foundation?

- LF AI is a LF umbrella foundation founded in March 2018 to support collaboration and open source innovation in the AI, ML, and DL domains.

- LF AI members are collaborating to create a neutral environment with an open governance for harmonization and acceleration of separate technical projects focused on AI, ML, and DL technologies.
Mission

Build and support an open AI community, and drive open source innovation in the AI, ML and DL domains by enabling collaboration and the creation of new opportunities for all the members of the community.
Members

Premier Members

General Members

Associate Members

The Associate membership category is limited to academic and nonprofit institutions respectively and requires approval by the Governing Board.
Goals

**Neutral Environment**

Provide an neutral environment and an open governance that fosters collaboration and cross-pollination

**Harmonization, Interoperability**

Increase efforts to harmonize AI/ML/DL open source projects and reduce fragmentation; increase the interoperability among projects

**Ethics and Fairness**

Become the home of open source AI efforts in ethics and fairness, and an advocate for these principles

**Data, Models Tools and Marketplace**

Host open source models with supporting tools; publish best practices; offer a marketplace for AI/ML/DL models

**Funding and Awareness**

Provide a funding model for key projects; increase awareness

---

A proven, vendor neutral not-for-profit foundation

TAC ML Workflow

Trusted AI Committee

Acumos marketplace

Funding model

LF AI Days

LF AI Summit

THE LINUX FOUNDATION
Projects

**Acumos**
Machine learning platform and marketplace for ML models.
Contributed by AT&T and Tech Mahindra

**Angel**
A high-performance distributed ML platform based on Parameter Server, running on YARN and Apache Spark.
Contributed by Tencent

**EDL**
Elastic Deep Learning
A framework designed to help DL cloud service providers to build cluster cloud services using different DL frameworks.
Contributed by Baidu

**HOROVOD**
Distributed training framework for TensorFlow, Keras, and PyTorch.
Contributed by Uber

**Pyro**
Pyro is a universal probabilistic programming language written in Python and supported by PyTorch on the backend. It enables flexible and expressive deep probabilistic modeling, unifying the best of modern deep learning and Bayesian modeling.
Contributed by Uber
# LF AI Project Maturities

<table>
<thead>
<tr>
<th>Incubation</th>
<th>Graduation</th>
<th>Emeritus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Projects join LF AI as incubation by meeting the requirements set forth in the “Project Process and Lifecycle Document”.</td>
<td>Incubation requirements.</td>
<td>Projects which the maintainers feel have reached or are nearing end-of-life. Emeritus projects have contributed to the ecosystem, but are not necessarily recommended for modern development as there may be more actively maintained choices.</td>
</tr>
<tr>
<td></td>
<td>Additional requirements as set in the “Project Process Lifecycle Document”.</td>
<td>TAC Vote + GB Vote.</td>
</tr>
</tbody>
</table>
LF AI Project Hosting Requirements

1. Use an approved OSI open source license
2. Be supported by a LF AI member
3. Fit within the mission and scope of LF AI
4. Allow neutral ownership of project assets such as a trademark, domain or GitHub account (the community can define rules and manage them)
5. Have a neutral governance that allows anyone to participate in the technical community, whether or not a financial member or supporter of the project
Universities Contributing to LF AI Projects

› Georgia Tech  
› University of British Columbia  
› Stanford University  
› MIT  
› Harvard University

› Oxford University  
› Penn University  
› NYU  
› Florida State University  
› University of Copenhagen
Work Groups and Ongoing Efforts
LF AI - ML Workflow
Reference workflow and implementation; integration and harmonization

Data Governance
- Data cleansing
- Data ingestion
- Data analysis & transformation
- Data validation
- Feature Engineering
- Data splitting

Trust Management
- Building a model
- Training optimization
- Model validation
- Training at scale
- Model

Security Management
- Deploying
- Serving
- Monitoring & Logging + Explainability
- Finetune & improvements

Dataflow and Workflow Orchestration
- Marketplace (AI Hub)
- Data consistency (versioning)

Data flow:
- Data prep: Data cleansing, Data ingestion, Data analysis & transformation, Data validation, Feature Engineering, Data splitting
- Model creation: Building a model, Training optimization, Model validation, Training at scale, Model
- Rollout: Deploying, Serving, Monitoring & Logging + Explainability, Finetune & improvements

THE LINUX FOUNDATION
LF AI
LF AI - Trusted AI Committee

› **Focus** - Policies, guidelines and tooling
› **Threads** - Fairness, robustness, explainability, and lineage
› **Documentation** - Content that describes the basic concepts and definitions in relation to Trusted AI in an effort to standardize the vocabulary and terminology
› **Badging Program** - Targeted for open source projects that meet the Trusted AI Ethics policies and guidelines as defined by LF AI
› **Training** - edX courses in collaboration with the LF training organization
Cross Collaboration Opportunities

**LF Edge**: Trusted AI is needed in edge devices, from driverless vehicles to smartphones to automated factories and farms. Industry fragmentation is a major challenge for edge development. LF Edge brings together projects across Internet of Things (IoT), cloud, and enterprises to increase unity across platforms, communities, and ecosystems. LF Edge fosters collaborations with end users, vendors, and developers to transform all aspects of edge technology and speed open source development.

**LF ODPi**: Data is at the heart of building open source trusted AI systems – and data governance is especially needed. ODPi provides one of the only vendor neutral, open source standards to enable best practices for data governance, connectivity, business intelligence, analytics, and AI systems.

**LF Energy**: The energy industry needs open source trusted AI across a wide range of business processes, from predicting demand to predictive maintenance of equipment and more. LF Energy provides a vendor-neutral, collaborative environment to “enable the electrification of everything to scale,” thereby transforming the world’s relationship to the important resource of energy.

**LF ONAP**: Trusted AI embedded in the network is a priority for the communications industry. The Open Network Automation Platform (ONAP) is ready to be infused with AI to enhance real-time, policy-driven orchestration and automation of physical and virtual network functions. Communication industry providers and developers can use open source to rapidly automate new services and support complete lifecycle management.

**LF CNCF**: The need for trusted AI in the cloud is the dominate way AI capabilities will be accessed for enterprise business processes. The Cloud Native Computing Foundation (CNCF) hosts critical components of the global technology infrastructure.
We are creating a sustainable open source AI ecosystem that makes it easier to create AI products and services using open source technologies
Joining LF AI Foundation is easy!

Please follow up with Ibrahim Haddad

- Web site: https://lfai.foundation
- Email: info@lfai.foundation
- Landscape: https://l.lfai.foundation
- GitHub: https://github.com/lfai
- Wiki: https://wiki.lfai.foundation
- Mailing lists: https://lists.lfai.foundation/groups